Functional association of CTCF with the insulator upstream of the *H19* gene is parent of origin-specific and methylation-sensitive

Chandrasekhar Kanduri*, Vinod Pant*, Dmitri Loukinov†, Elena Pugacheva†, Chen-Feng Qi†, Alan Wolffe‡, Rolf Ohlsson* and Victor V. Lobanenkov‡

In mammals, a subset of genes inherit gametic marks that establish parent of origin-dependent expression patterns in the soma ([1] and references therein). The currently most extensively studied examples of this phenomenon, termed genomic imprinting, are the physically linked *Igf2* (insulin-like growth factor II) and *H19* genes, which are expressed mono-allelically from opposite parental alleles [1,2]. The repressed status of the maternal *Igf2* allele is due to *cis* elements that prevent the *H19* enhancers [3] from accessing the *Igf2* promoters on the maternal chromosome [4,5].

A differentially methylated domain (DMD) in the 5′ flank of *H19* is maintained paternally methylated and maternally unmethylated [6,7]. We show here by gel-shift and chromatin immunopurification analyses that binding of the highly conserved multivalent factor CTCF ([8,9] and references therein) to the *H19* DMD is methylation-sensitive and parent of origin-dependent. Selectively mutating CTCF-contacting nucleotides, which were identified by methylation interference within the extended binding sites initially revealed by nuclease footprinting, abrogated the *H19* DMD enhancer-blocking property. These observations suggest that molecular mechanisms of genomic imprinting may use an unusual ability of CTCF to interact with a diverse spectrum of variant target sites, some of which include CpGs that are responsible for methylation-sensitive CTCF binding *in vitro* and *in vivo*.

Addresses: *Department of Development and Genetics, Uppsala University, S-75236 Uppsala, Sweden. †Molecular Pathology Section, Laboratory of Immunopathology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. ‡Sangamo Biosciences Inc., Point Richmond Tech Center, 501 Canal Blvd, Suite A100, Richmond, California 94804, USA.

Correspondence: Victor V. Lobanenkov, Rolf Ohlsson
E-mail: vlobanenkov@niaid.nih.gov
Rolf.Ohlsson@ebc.uu.se

Received: **17 May 2000**
Revised: **5 June 2000**
Accepted: **6 June 2000**

Published: **30 June 2000**

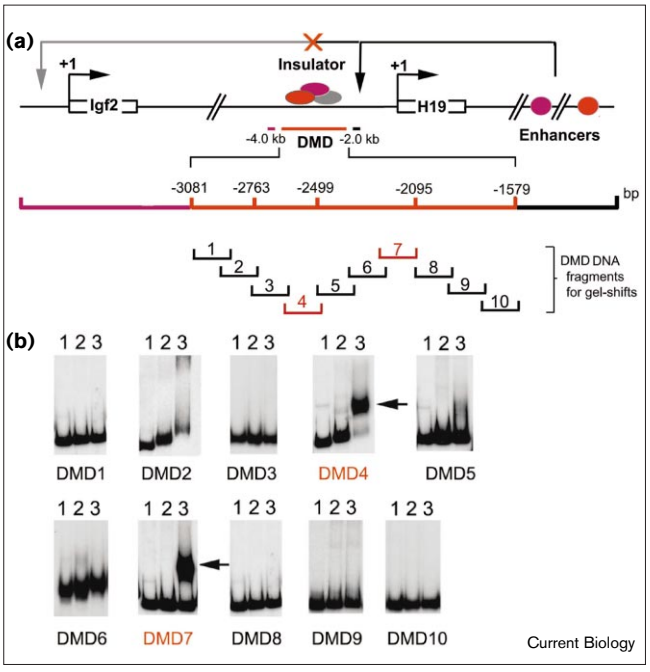
Current Biology 2000, 10:853–856

0960-9822/00/$ – see front matter

Results and discussion

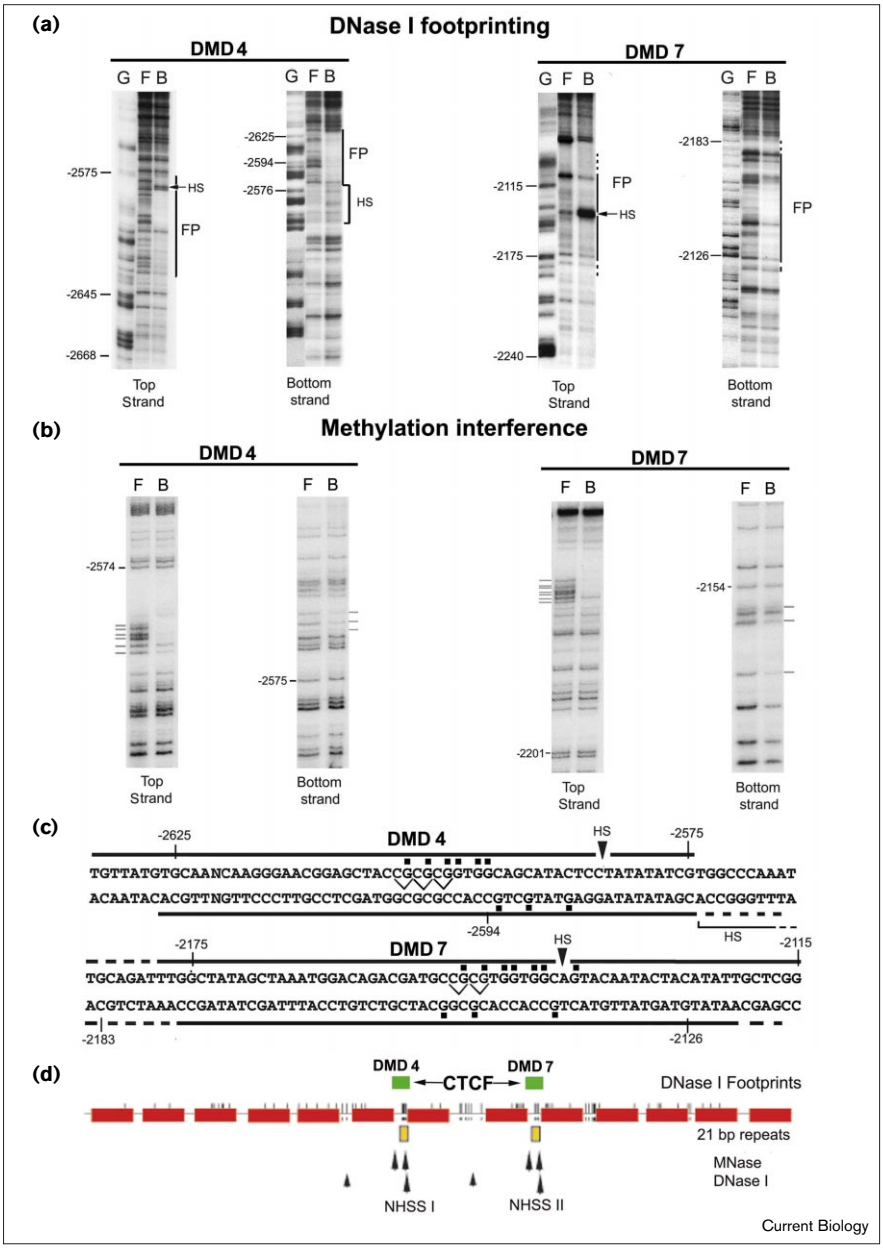
The chromatin conformation of the *H19* DMD displays prominent nuclease-hypersensitive sites (NHSSs) at the 21 bp repeat conserved in mammals ([5] and references therein). The repeat consensus has some sequence similarity to the CTCF protein target sequences ([9] and references therein). The highly conserved and ubiquitously expressed CTCF protein has 11 Zn-fingers (ZFs). It is an unusual multivalent factor capable of binding to remarkably different and ~50–60 bp long CTS sequences by utilizing different sets of ZFs ([8,9] and references therein). CTSs mediate promoter repression or activation and help to create hormone-responsive silencers ([8,9] and references therein). In addition, core sequences of a number of diverse chromatin insulators in vertebrates have been narrowed down to CTSs [10].

Figure 1



Systematic screening of the *H19* DMD for CTCF-binding sites. **(a)** Schematic map of the mouse *Igf2–H19* genomic region with functionally relevant regulatory elements and arrangement of consecutive, overlapping DNA fragments, which were each labeled at the 5′ end. Arrows indicate the activation of each promoter by the *H19* enhancers. Numbering is relative to the +1 *H19* transcription start. **(b)** Gel-shift assays with the DMD1 to DMD10 DNA fragments and the 11 ZF domain of CTCF synthesized from the pCITE4a-11ZF vector [9]. Lanes 1, 2, and 3 on each panel correspond to gel-shift reactions with no protein, with the negative luciferase protein control, and with CTCF, respectively. Fragments producing shifted complexes (indicated by arrows) are in red.

Figure 2



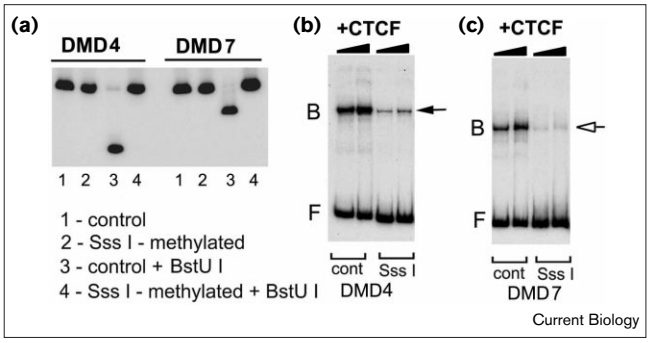
Characterization of the DMD4 and DMD7 CTCF-binding sequences. **(a)** DNase I footprinting and **(b)** DMS-methylation interference assays were carried out with full-length CTCF. G, the Maxam–Gilbert sequencing G-ladders; F and B, free and CTCF-bound DNA probes, respectively; FP, footprint regions protected from nuclease attack; HS, DNase I hypersensitive sites induced upon CTCF binding. In (b), guanines that cannot be modified by DMS without losing contact with CTCF are indicated by green bars. **(c)** Summary of the results shown in (a,b). Critical contact G residues are indicated by filled rectangles; on each strand, DNA sequences protected by CTCF from DNase1 are underlined or overlined. The CpG pairs (*BstU*I sites) that include dGs critical for CTCF recognition are indicated by angle brackets. **(d)** The CTCF chromatin map of the maternally derived *H19* DMD allele. Red rectangles depict estimated nucleosome positions, whereas the vertical bars identify CpG dinucleotides [5]. The 21 bp conserved repeats are indicated by yellow boxes, and CTCF-produced DNase I footprints are shown by green boxes. The numbers indicate nucleotide positions relative to the +1 transcriptional start site of the *H19* gene.

To test for a link between CTCF and the *H19* DMD function, we focused on the second two thirds of the *H19* DMD where we could persistently document presence of the NHSSs upstream of NHSS I in mouse fetal liver [5]. We carried out systematic electrophoretic mobility shift assays (EMSA) for CTSs in this *H19* region and detected two new CTSs, termed DMD4 and DMD7 (Figure 1). To define exactly which sequences are occupied by CTCF, and to identify guanines within these sequences that cannot be modified without losing CTCF binding, we performed DNase I footprinting and methylation interference assays. The results of these experiments, shown in Figure 2, led us to the following conclusions: first, the positions of the DMD4 and DMD7 CTSs correspond precisely to NHSS I and NHSS II, respectively; second, in each recognition sequence, CTCF protects ~60 bp of both DNA strands from nuclease attack; third, inside of each DMD CTS, CTCF induces DNase-hypersensitivity subsites on the top GC-rich strand; and fourth, in both DMD4 and DMD7 CTSs, the identical CGCG(T/G)GGTG-GCAG core motif provides major contact bases for recognition by CTCF. DMD4 and DMD7 core CTSs contain three and two CpG pairs, respectively, which are methylated *in vivo* when derived from father [6,7]. Although the

DMD4 and DMD7 CTSs are similar one to another, they are different from several other previously identified CTSs [8–10], suggesting a unique contribution of interacting ZFs. Indeed, EMSAs with the proteins containing 12 serial truncations from either end of the 11 ZF domain [9] fully supported this idea (see Supplementary material).

To test whether methylation might interfere with CTCF binding, we modified DMD4 and DMD7 CTS fragments with the *Sss*I methylase and compared them with the unmodified CTS for CTCF binding. Our EMSAs revealed a marked preference of CTCF for the unmethylated DMD sites (Figure 3). We also quantitatively estimated the effect of CpG methylation on the affinity of CTCF binding to each DMD CTS by surface plasmon resonance using the BIACORE X device. It appeared, quite unexpectedly, that the best-fit model for CTCF-DNA interaction is the two-stage reaction, with an intermediate conformational change resulting in formation of stable non-dissociating complexes with an apparent affinity constant in the range of 10^{11} to 10^{13} M^{-1}. In contrast, CTCF binding to the methylated DMD4 and DMD7 sites was at least a 1,000-fold lower in affinity (~10^{8} M^{-1}), and no stable complexes with methylated probes were detected (data not shown). CTCF affinity to methylated DMDs was still high enough to detect some residual binding by EMSAs not designed to measure the on-and-off binding rates (see Supplementary material).

Figure 3



CTCF binding to *H19* DMD is methylation-sensitive. **(a)** 5′-end-labeled control and *Sss*I methylase-treated DMD4 and DMD7 fragments were digested with methylation-sensitive *BstU*I and analyzed on polyacrylamide gels to verify efficiency of *in vitro* methylation. **(b,c)** Control unmethylated (cont) or *Sss*I-methylated DMD4 and DMD7 DNA fragments were analyzed by gel-shift assays with increasing amounts of CTCF as indicated at the top of each panel. F, free probe; B, CTCF-bound probe.

To determine whether the *H19* DMD CTS sequences display the chromatin-insulator-like activity, we generated

Figure 4

The CTSs in *H19* DMD mediate the enhancer-blocking properties of the *H19* 5′ flank. **(a)** EMSA shows that point mutations within the DMD (see text and diagram below) destroy the CTCF recognition elements. F, free probe; B, bound probe; control is luciferase. **(b)** Enhancer-blocking assay in JEG-3 cells to assess effects of targeted mutations in the CTSs. The assay determines the ability of the wild-type or mutant *H19* DMD to prevent the SV40 enhancer from communicating with the promoter of the *H19* reporter gene (see diagram below). GAP, glyceraldehyde 3-phosphate dehydrogenase. **(c)** Quantitation of enhancer-blocking assay normalized to DNA input and episome copy number. The SV40 enhancer-driven expression of the pREPH19A construct was, for convenience, assigned a value of 100 while all other samples were normalized relative to this value. The mean deviation of at least three different experiments is indicated for each vector construct, unless the differences were too insignificant to show.

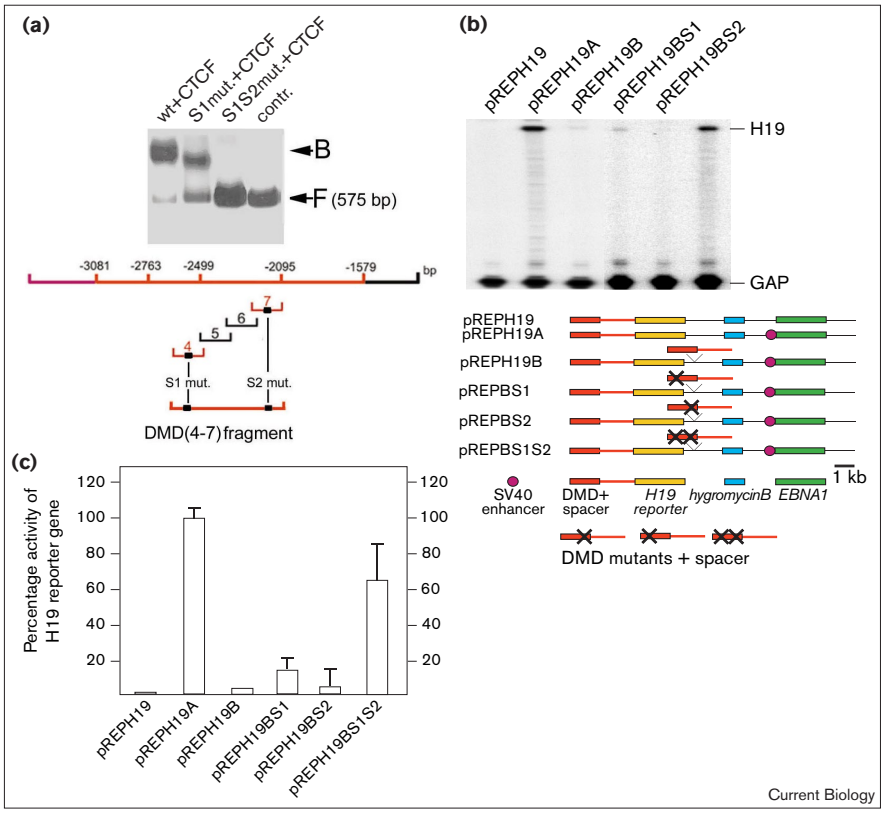
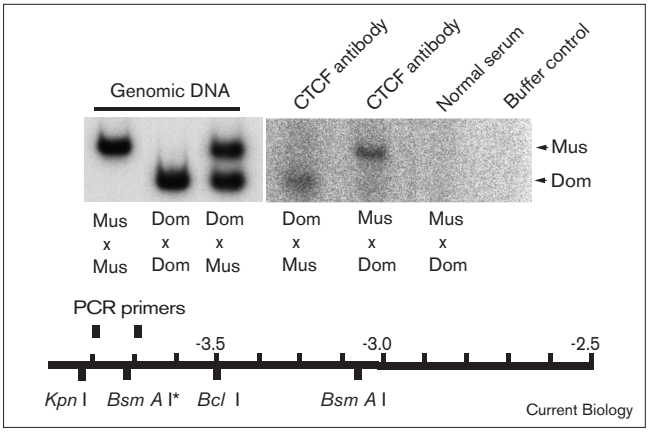


Figure 5



Parent of origin-specific association of CTCF with the chromatin of the *H19* 5′ flank. Formaldehyde-cross-linked DNA, which was derived from fetal liver of reciprocal intraspecific hybrid crosses of *M. m. domesticus* (D) and *M. m. musculus* (M), was immunopurified with an antibody to CTCF followed by PCR amplificaton. The PCR primers spanned a polymorphic *BsmA*I site which was specific for the M allele. The M × D or D × M crosses are indicated in the order mother–father in all instances.

point mutations that eliminate CTCF interaction with the DMD4 and DMD7 sites. Figure 4a shows that changing the sequence GTGG to ATAT in either of the DMD CTSs (see Figure 2c) destroyed their interaction with CTCF. Next, we addressed whether the enhancer-blocking properties of the *H19* DMD depend on the CTSs in an episomal-based assay [5]. As could be expected, the targeted disruption of CTCF-DMD interaction at both CTS counteracted most of the enhancer-blocking properties of the *H19* 5′ flank (Figure 4b,c).

To examine an *in vivo* link between CTCF and the *H19* 5′ flank, we analyzed the distribution of CTCF in formaldehyde-crosslinked chromatin of fetal livers of reciprocal *M. musculus musculus* (M) × *M. musculus domesticus* (D) intraspecific hybrid crosses (see Figure 5 legend). CTCF-immunoprecipitated DNA was analyzed by a PCR assay, which allowed the discrimination of the parental alleles of the *H19* 5′ flank by exploiting a polymorphic restriction site, *BsmA*I [5]. Figure 5 shows that only the maternally inherited allele (the M allele in the M × D cross) is specifically captured by the CTCF antibody. When the opposite cross (D × M) was examined, the D allele was preferentially amplified. Given that the average length of the sonicated DNA fragments was 2–3 kb, most if not all of the potential CTCF-binding sites scattered within the DMD of the *H19* 5′ flank are likely to be covered in this assay. We conclude that the CTCF interaction with the *H19* 5′ flank is parent of origin-specific.

We have shown here that CTCF is both structurally and functionally an integral part of the *H19* DMD chromatin conformation and that its parent of origin-dependent interaction with the *H19* insulator is likely to involve its ability to read the methylation mark. It is tempting to speculate that function of CTCF as a candidate tumor suppressor gene at chromosome segment 16q22.1, where the predicted third Wilms' tumor gene (WT3) is localized, and frequent loss of *Igf2/H19* imprinting in these tumors ([8] and references therein) may be causally linked. Although our data implicate CTCF in manifesting the repressed status of the maternal *Igf2* allele in the soma, this remains to be demonstrated in targeted deletion mouse models.

Supplementary material

Supplementary material including figures showing the combinations of ZFs required to bind different DMDs and the methylation-sensitivity of CTCF binding to DMDs, and additional discussion and methodological details, is available at http://current-biology.com/supmat/supmatin.htm.

Acknowledgements

We acknowledge valuable advice from Gary Franklin, Joe Breen, and Herbert Morse. This work was supported by the Intramural NIAID NIH Research Funding to V.V.L., and by the Natural Science Research Council (NFR), the Swedish Cancer Research Foundation (RmC/CF), and the Swedish Pediatric Cancer Foundation (BCF) to R.O.

References

1. Ohlsson, R. (ed): **Genomic imprinting, an interdisciplinary approach.** In *Results and Problems in Cell Differentiation,* vol 25. Edited by Hennig W, Nover L, Scheer U. Berlin, Heidelberg, New York: Springer; 1999.
2. Bartolomei MS, Tilghman SM: **Genomic imprinting in mammals.** *Annu Rev Genet* 1997, **31:**493-525.
3. Leighton P, Saam J, Ingram R, Stewart C, Tilghman S: **An enhancer deletion affects both *H19* and *Igf2* expression.** *Genes Dev* 1995, **9:**2079-2089.
4. Thorvaldsen JL, Duran KL, Bartolomei MS: **Deletion of the *H19* differentially methylated domain results in loss of imprinted expression of *H19* and *Igf2*.** *Genes Dev* 1998, **12:**3693-3702.
5. Kanduri C, Holmgren C, Pilartz M, Franklin G, Ullerås E, Kanduri M, *et al.*: **The 5′-flank of the murine *H19* gene in an unusual chromatin conformation unidirectionally blocks enhancer-promoter communication.** *Curr Biol* 2000, **10:**449-457.
6. Olek A, Walter J: **The pre-implantation ontogeny of the *H19* methylation imprint.** *Nat Genet* 1997, **17:**275-276.
7. Tremblay K, Duran K, Bartolomei MS: **A 5′ 2-kilobase-pair region of the imprinted mouse *H19* gene exhibits exclusive paternal methylation throughout development.** *Mol Cell Biol* 1997, **17:**4322-4329.
8. Filippova GN, Lindblom A, Meincke LJ, Klenova EM, Neiman PE, Collins SJ, *et al.*: **A widely expressed transcription factor with multiple DNA sequence specificity, CTCF, is localized at chromosome segment 16q22.1 within one of the smallest regions of overlap for common deletions in breast and prostate cancers.** *Genes Chromosomes Cancer* 1998, **22:**26-36.
9. Awad TA, Bigler J, Ulmer JE, Hu YJ, Moore JM, Lutz M, *et al.*: **Negative transcriptional regulation mediated by thyroid hormone response element 144 requires binding of the multivalent factor CTCF to a novel target DNA sequence.** *J Biol Chem* 1999, **274:**27092-27098.
10. Bell A, West A, Felsenfeld G: **The protein CTCF is required for the enhancer blocking activity of vertebrate insulators.** *Cell* 1999, **98:**387-396.